

# Report on Lorentz Workshop

## Capturing Phylogenetic Algorithms for Linguistics

Marian Klammer, Harald Hammarström, Devdatt Dubhashi  
Gerhard Jäger, Andrew Meade and Russell Gray

November 9, 2015

### Science

*A concise version of the description and aims of the workshop, including what were the most important scientific questions motivating the workshop (about 5 lines).*

The past decade has seen an explosion in the availability of large datasets of both structural and lexical linguistic data that can now be exploited computationally to explore linguistic history and linguistic universals. The workshop aimed to bring together bioinformaticians, mathematicians and linguists to present work on how to modeling of linguistic data quantitatively, algorithms for inferences on linguistic data, case studies of specific families/regions and synergies between linguistic and non-linguistic data.

*Is a tangible outcome of the workshop expected? If so, please mention – even if it is, as yet, at the level of intention.*

There will be a proceedings and a state-of-the-art survey paper by the organising team. Some members of the organizing team were able to meet and get an impression on the scientific capabilities of potential candidates for jobs. Several possibilities for joint project applications were discussed.

*Where there any developments which could, already, be termed a (beginning) scientific breakthrough? If yes, please tell about it shortly.*

Many interesting problems were discussed and several subproblems were solved but a “breakthrough” is a too strong word.

*Did you, or to your knowledge any of the participants, experience notable “Aha” moments (for instance, separate scientific communities realizing that they have significantly more in common than they had thought)?*

Many participants reported revelations after John Huelsenbeck’s very accessible tutorial on methods that involve advanced mathematical knowledge (also continued in a more interactive manner during the discussion sessions).

Jotun Hein was very active in the discussion sessions and explained the progress on algorithms for string comparison developed in Bioinformatics, which turns out to be extremely similar to the problem(s) on word (form) comparison in linguistics.

Eva Boon presented the state-of-the-art of data resources and infrastructure in Bioinformatics. The world of Bioinformatics databases and Linguistics databases turn out to be almost completely parallel, except that the Bioinformatics world is far larger.

## Organization/Format

*How did you experience the format of the workshop (the structure of the program, lectures vs discussion time etc.)? Did you try something new (different kind of discussions for instance)? If so, how did it work out? Would you do it again or advise it to others?*

The attendants consumed the (full) program lasting five days surprisingly well, no doubt due to the generally high level of presentations and participation.

We attempted discussion sessions in the form of state-of-the-art sessions with mixed results, and one session of joint discussion which especially the junior participants found very rewarding.

Many potential attendants outside the main core of participants were interesting in coming ad-hoc for a day or two. The Lorentz Centre were helpful in accommodating this, and, in retrospect, we should have been clearer (towards these potential attendants) on the possibilities to do so.

*Other comments, suggestions and/or criticism for the Lorentz Center, the scientific advisory boards and/or future organizers.*

The long turn-around with reviews, our response to reviews, the final approval from the Lorentz Centre significantly delayed us. Ideally, the process should have started earlier or be slimmed down to be quicker. Many of the comments from the reviewers were difficult to address, i.e., those that called for more detailed information than the two pages allotted admit and those that were general (rather than addressing the relation between the aims of the workshop and the stated plan of achieving them). Possibly the reviewers can be given some more direction at reviewing the relation between the goals and the plan. The reviews also contained some very good suggestions for improvement.

The webpage capabilities seem unnecessarily restrictive since they did not allow more than two bullet points and could not list the invited speakers without a click to open a pdf containing only this list.

*If in the coming years, papers are published which you believe owe significantly to your workshop, please notify us. We find that highly gratifying fruits of our work and it is naturally desirable for fundraising.*

Will do.